**Noname manuscript No.**
(will be inserted by the editor)

# Successes and challenges in developing a hybrid approach to sentiment analysis

**Orestes Appel · Francisco Chiclana ·**
**Jenny Carter · Hamido Fujita**

Received: date / Accepted: date

**Abstract** This article covers some success and learning experiences attained during the developing of a *hybrid approach* to Sentiment Analysis (SA) based on a Sentiment Lexicon, Semantic Rules, Negation Handling, Ambiguity Management and Linguistic Variables. The proposed hybrid method is presented and applied to two selected datasets: Movie Review and Sentiment Twitter datasets. The achieved results are compared against those obtained when Naïve Bayes (NB) and Maximum Entropy (ME) supervised machine learning classification methods are used for the same datasets. The proposed hybrid system attained higher *accuracy* and *precision* scores than NB and ME, which shows its superiority when applied to the SA problem at the sentence level. Finally, an alternative strategy to calculating the orientation polarity and polarity intensity in one step instead of the two steps method used in the hybrid approach is explored. The analysis of the yielded mixed results achieved with this alternative approach shows its potential as an aid in the computation of semantic orientations and produced some lessons learnt in developing a more effective mechanism to calculating the orientation polarity and polarity intensity.

Orestes Appel
Bissett School of Business, Mount Royal University, Calgary, Alberta, Canada.
Centre for Computational Intelligence (CCI), Faculty of Technology, De Montfort University, Leicester, UK.
E-mail: oappel@mtroyal.ca, orestes.appel@my365.dmu.ac.uk

Francisco Chiclana and Jenny Carter
Centre for Computational Intelligence (CCI), Faculty of Technology, De Montfort University, Leicester, UK.
E-mail: chiclana@dmu.ac.uk, jennyc@dmu.ac.uk

Hamido Fujita
Iwate Prefectural University (IPU), Takizawa, Iwate, Japan
E-mail: kissam@iwate-pu.ac.jp

**Keywords** Sentiment Analysis · Fuzzy Sets · Semantic Rules · Natural Language Processing · Computational Linguistic · Uninorms · SentiWordNet · Computing with Sentiments

## 1 Introduction

Sentiment Analysis (SA) has been an area of intensive research during the last few years. The data volumes nowadays available are too voluminous and complex for human consumption. As a consequence, the need for automated computer-based capabilities that could tell us whether sentences or tweets are carrying opinions or factual information, is of great importance. In addition to discriminating positive sentences from negatives, there is also room for quantifying the degree of positiveness or negativeness of a given piece of information. Generaly, SA is performed at specific levels, such as feature/aspect level, sentence level, document level, etc. The focus of this article is SA at the sentence level. A complete review of the evolution of the Sentiment Analysis field can be obtained from reading [1] and [2].

Common approaches to the SA problem belong either to the Supervised or Unsupervised Machine Learning categories. However, fuzzy sets ability to deal with vagueness and uncertainty have also been considered well equipped to be applied to model SA related problems. Indeed, Dzogang et al. in [7] provided examples of appraisal-based approaches that make use of graduality using fuzzy inference and fuzzy aggregation as tools for processing affective mechanisms ambiguity and imprecision. Additionally, Bing Liu in [15] claims that it is possible that researchers have relied too much in *machine learning*. These arguments constitute reasonable grounds for alternative approaches to the SA problem to be pursued and worth investigating. Indeed, it can be hypothesised that a combination of different techniques could be more effective at addressing the challenges of SA than specific techniques in isolation. A hybrid methodology approach to the SA problem at the sentence level based on: (i) the proven success of the use of NLP techniques (e.g. smart parsing, tokenisation, etc.); (ii) the extraordinary contribution that a solid opinion lexicon and semantic rules could have in polarity determination; and (iii) the concept of *graduality* expressed through fuzzy sets, in order to compute the intensity of a given sentiment polarity, is investigated and applied to two selected datasets: Movie Review and Sentiment Twitter datasets. The achieved results are compared against those obtained when Naïve Bayes (NB) and Maximum Entropy (ME) supervised machine learning classification methods are applied to the same datasets. As it will be shown later in the article, the proposed hybrid system attained higher *accuracy* and *precision* scores than NB and ME, which shows its superiority when applied to the SA problem at the sentence level. Additionally, the possibility of calculating both, polarity and polarity intensity in only one step, rather than the two steps used in the mentioned hybrid approach, is investigated. This alternative polarity quantification approach is based on the work from a number of authors, among them Fodor's work [9] on uninorms and extensions of the cross-ratio uninorm with neutral element $e = 0.5$ to the case of having neutral element any value $e$ in interval $[0, 1]$ or $[-1, 1]$. The alternative strategy to calculating the orientation polarity and polarity intensity in one step when applied to the same datasets yields mixed results that shows its potential as an aid in the computation of semantic orientations. Some lessons can be learnt from such experimentation in developing a more effective mechanism to calculating the orientation polarity and polarity intensity, which is elaborated further later in the article.

Recently, other researchers have explored as well the possibility of using fuzzy sets in sentiment analysis, like the case of the article by Loia and Senatore [16], where they explore the alternative of representing sentiments inspired on the Minsky's conception of emotions (in their work, sentiments and emotions are modelled as fuzzy sets).

Another school of thought in sentiment analysis is to use concepts of emotions derived from Psychology to discover concealed sentiments in language, like the work of Brenga et al. [3], where the authors introduced a framework for detecting sentiment and emotion from text based on an affective model known as Hourglass of Emotions (a variant of Plutchik's wheel of emotions).

The rest of this article is structured as follows: Section 2 covers the proposed Hybrid Classification Method for SA; in Section 3 a comparison of the experimental results obtained when applying the Hybrid Classification Method, NB and ME to two selected datasets, Movie Review and Sentiment Twitter datasets, is presented. An alternative method that assisted us in confirming the efficiencies of our proposed Hybrid Classification Method for SA will be studied in Section 4. Finally, conclusions and some ideas that could be constructed as further research work will be presented in Section 5.

## 2 The Proposed Hybrid Approach to Sentiment Analysis

The main components of the proposed SA method are:

- The Sentiment Lexicon
- The Semantic Rules
- The Negation Handling Process
- The Ambiguity Management Process
- The Linguistic Variables

The next sub-sections will fully describe these aspects.

### 2.1 The Sentiment Lexicon

In generating our Sentiment Lexicon the following approach has been used:

1. The *opinion-conveying-words* ('positive meaning words' and 'negative meaning words') of the Opinion Lexicon used by Prof. Bing Liu et al. in [12] are

utilised. *Only* the terms proven capable of delivering opinions elements of Part-of-Speech (PoS) are considered [10,11,13,28] : nouns; verbs; adjectives and adverbs.

2. SentiWordNet [8] is used to extract *polarity* for words carrying opinion sense as well as their associated PoS tags.
3. Both elements in (1) and (2) above are combined. Words in the original Liu's opinion lexicons are substituted for equivalent Synsets in SentiWordNet. As a consequence, *positive* and *negative* scores as well as PoS tags are added to the existing words in Liu's lexicon. All of the above contribute to the generation of an *improved* Sentiment Lexicon.

### 2.2 Semantic Rules (SR)

A number of authors [18,26,29] have mentioned that the final outcome of a classification algorithm could be affected by *negation* and the use of specific *PoS particles* such as but, despite, unless. Hence, a proper rule strategy is required in order to deal with different PoS particles that could play a key role in sentence semantic classification. Through time, researchers have improved the quality of these semantic rules [29]. During our research we utilised some of these aforementioned rules, and incorporated others with the objective of managing two particular PoS particles that were not included in the original set of rules given in [29]: **while** and **however**. As such, the resulting semantic rules are those displayed in Table 1. Notice that rules given in [29] not utilised in our research are clearly identify in Table 1.

### 2.3 Negation Handling

"Sentiment words behave very differently when under the semantic scope of negation" [24]. The complex nature of *negation* precludes the possibility of counting with *a priori* rules. The technique embraced in our research utilises *Regular Expressions* [6,21] and can manage long-distance negation effects. Negation processing is performed at two different times:

– Syntactic Analysis time: the regular expression mechanism mentioned above determines the scope of the negation and tag the words in the scope of the negation accordingly. Let us see a specific sentence as an example:

  'I don't think I will enjoy it: it might be too spicy'.

  As per the negation handling technique just mentioned, all words between the negation particle 'don't' and the colon ':' will be tagged with the suffix '**_NEG**', clearly defining the scope of the negation. Words after the colon ':' would not be tagged at all. Doing this at the syntactic analysis time serves two purposes: (i) first of all, it saves us time as the scope of negation is defined early on, and if a polarity inversion is required, it can be

**Table 1** Semantic Rules for proposed Hybrid System

| **Rules** | **Semantic Rules** | **Example** |
|---|---|---|
| R1 | Polarity (not $var_k$) = -Polarity ($var_k$) | 'not *bad*.' |
| R2 (not used) | Polarity ($NP_1$ of $NP_2$) = *Compose* ($NP_1$, $NP_2$) | '*Lack* of crime in rural areas.' |
| R3 | Polarity ($NP_1$ $VP_1$) = *Compose* ($NP_1$, $VP_1$) | '*Crime* has decreased.' |
| R4 (not used) | Polarity ($NP_1$ be $ADJ$) = *Compose* ($ADJ$, $NP_1$) | '*Damage* is minimal.' |
| R5 (not used) | Polarity ($NP_1$ of $VP_1$) = *Compose* ($NP_1$, $VP_1$) | '*Lack* of killing in rural areas.' |
| R6 | Polarity ($ADJ$ to $VP_1$) = *Compose* ($ADJ$, $VP_1$) | '*Unlikely* to destroy the planet.' |
| R7 | Polarity ($VP_1$ $NP_1$) = *Compose* ($VP_1$, $NP_1$) | '*Destroyed* terrorism.' |
| R8 (not used) | Polarity ($VP_1$ to $VP_2$) = *Compose* ($VP_1$, $VP_2$) | '*Refused* to deceive the man.' |
| R9 (not used) | Polarity ($ADJ$ as $NP$) = $1_{(Polarity(NP=0))} \cdot$ Polarity($ADJ$) + $1_{(Polarity(NP\neq 0))} \cdot$ Polarity($NP$) | 'As *ugly* as a rock.' |
| R10 | Polarity (not as $ADJ$ as $NP$) = -Polarity ($ADJ$) | 'That wasn't as *bad* as the original.' |
| R11 | If sentence contains "but", disregard all previous sentiment and only take the sentiment of the part after "but". | 'And I've never liked that director, *but* I loved this movie.' |
| R12 | If sentence contains "despite", only take the sentiment of the part before "despite". | 'I love the movie, *despite* the fact that I hate that director.' |
| R13 | If sentence contains "unless" followed by a negative clause, disregard the "unless" clause. | 'Everyone likes the video *unless* he is a sociopath.' |
| R14 (New) | If sentence contains "while", disregard the sentence following the 'while' and take the sentiment only of the sentence that follows the one after the 'while'. | '*While* they did their best, the team played a horrible game.' |
| R15 (New) | If sentence contains "however", disregard the sentence *before* 'however' and take only the sentiment of sentence *after* 'however'. | 'The film counted with good actors. *However*, the plot was very poor.' |

done at tokenization time during syntactic analysis; (ii) secondly, if a part of a sentence is identified as one that will not contribute to the final semantic orientation, then such part of the sentence can be discarded at this point, minimising the effort required at sentiment computing time. When the time comes to compute semantic orientation, terms tagged with the particle '_NEG' will be altered in two ways: (i) the polarity label of the affected word will be inverted (Positive converted to Negative, or Negative to Positive), and the polarity scores will be inverted as well (i.e. a Positive Score = 0.7, and a Negative Score = 0.3 would be transformed into a Positive Score = 0.3, and a Negative Score = 0.7). Hence, the preparation for *semantic analysis* is completed and when the time arrives to determine the polarity intensity via fuzzy sets, the polarity scores (Positive/Negative) of the involved words have been already updated reflecting the real orienta-

**Table 2** Compose functions referenced in Table 1

| **Compose Functions** | **Algorithms** |
|---|---|
| Compose1 (arg1, arg2) | 1. Return **-**Polarity(*arg2*) if *arg1* is negation. |
| | 2. Return Polarity(*arg1*) if (Polarity(*arg1*) = Polarity(*arg2*). |
| | 3. Otherwise, return the majority term polarity in *arg1* and *arg2*. |
| Compose2 (arg1, arg2) | 1. Return Polarity(*arg2*) if *arg1* is negative and *arg2* is not neutral. |
| | 2. Return **-**1 if *arg1* is negative and *arg2* is neutral. |
| | 3. Return Polarity(arg2) if *arg1* is positive and *arg2* is not neutral. |
| | 4. Return 2*****Polarity(arg1) if Polarity(*arg1*) = Polarity(*arg2*). |
| | 5. Return Polarity(*arg1*) **+** Polarity(*arg2*) if *arg1* is positive and *arg2* is neutral. |
| | 6. Return Polarity(*arg1*) **+** Polarity(*arg2*) if *arg2* is positive and *arg1* is neutral. |
| | 7. Otherwise, return **0**. |

tion of the words after having solved the negation aspect.

- Semantic Rules Application time: the Semantic Rules dealing with negation will be call into action to process particles previously tagged (step above) as in-scope of negation (see R1 above in Table 1, Sub-section 2.2).

### 2.4 Ambiguity Management

Ambiguity is a common occurrence in many languages, including English, as the same word can play different functions assuming different part-of-speech roles. In our research ambiguity is managed through the part-of-speech tagging and parsing processes. When a sentence is processed and the syntactic analysis is completed, including negation, only relevant words, those capable of carrying a sentiment, are left in the group to be studied. At parsing time, each word get assigned a tag corresponding to an existing list of possible parts of speech. In the sentiment lexicon each word has associated a part-of-speech tag, which implies that some words may occur more than once in the lexicon if they can take on different part-of-speech roles (verb, noun, adverb, adjective, etc.). That way, it would be possible, most of the times, to determine that role played by a given word once the parse tree has been built. There will always be exceptions, as ambiguity management is an active area of research in NLP. More specifically, let us look at an example:

*Example 1* Flies like a flower.:

- Flies: is it a noun or verb?
- like: is it a preposition, an adverb, a conjunction, a noun or a verb?
- a: is it an article, a noun, or a preposition?

– flower: is it a noun or a verb?

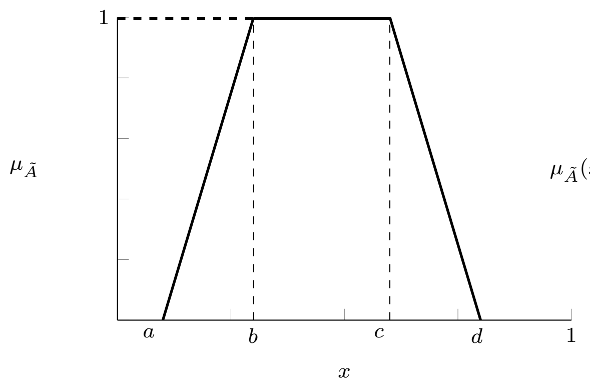
By using part-of-speech tagging at syntactic analysis time, and assigning to every particle the proper part-of-speech tag, the ambiguity problem is addressed. The above sentence would look like this, once parsed and tagged:

( (flies VB) (like ADV) (a AT) (flower NN) ), where

VB = Verb
ADV = Adverb
NN = Noun
AT = Article

## 2.5 Linguistic Variables and Fuzzy Sets

The classification of objects without the need to actually measuring them is a natural ability humans have developed thanks to the use of natural languages. When we say that an object is *slightly* large or *very* large, we all understand the meaning. Nevertheless, we have not measured at all the object we are referring to. In these cases, though, a constraint to keep in mind is the number of categories that a subject can effectively maintain, which according to Miller's study [17] is *7 plus or minus 2*. Hence, when devising the *linguistic variables* in our proposed hybrid SA model, we have chosen five labels (7 minus 2) symmetrically distributed in the unit interval, represented using the following 4-tuple trapezoidal membership functions (MFs) [22], as generally shown in Fig. 1:




$$\mu_{\tilde{A}}(x) = \begin{cases} 0 & \text{if} \quad x \leq a; \\ \frac{x-a}{b-a} & \text{if} \quad a \leq x \leq b; \\ 1 & \text{if} \quad b \leq x \leq c; \\ \frac{d-x}{d-c} & \text{if} \quad c \leq x \leq d; \\ 0 & \text{if} \quad d \leq x. \end{cases}$$

**Fig. 1** Trapezoidal membership function

– MF (Poor): (0, 0, 0.050, 0.150)
– MF (Slight): (0.050, 0.150, 0.250, 0.350)
– MF (Moderate): (0.250, 0.350, 0.650, 0.750)
– MF (Very): (0.650, 0.750, 0.850, 0.950)
– MF (Most): (0.850, 0.950, 1, 1)

Perceptions are commonly utilised. According to Zadeh [31]: "reflecting the bounded ability of the human brain to resolve detail, perceptions are intrinsically imprecise. In more concrete terms, perceptions are f-granular, meaning that (1) the boundaries of perceived

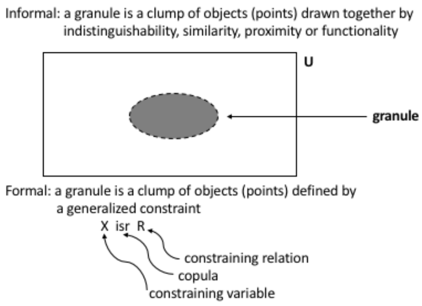



**Fig. 2** The Concept of a Granule as presented by Zadeh

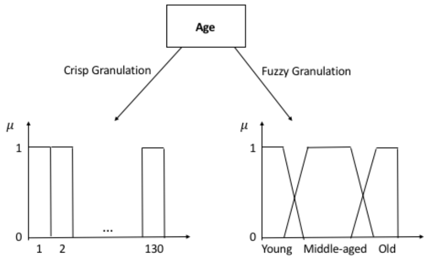



**Fig. 3** Crisp Granulation and Fuzzy Granulation as introduced by Zadeh

classes are unsharp and (2) the values of attributes are granulated, with a granule being a clump of values (points, objects) drawn together by indistinguishability, similarity, proximity, and function" (Fig. 2 re-illustrates the original graphic published in [32]). In [32], Zadeh continues, by saying that "a granule may be crisp or fuzzy, depending on whether its boundaries are or are not sharply defined. For example, age may be granulated crisply into years and granulated fuzzily into fuzzy intervals labeled very young, middle-aged, old and very old." Figure 3 re-illustrates the graphical representation of the latter idea as originally presented in [31].

Notice that the *intensity* or *degree of polarity* with which the grade of positivity or negativity of a sentence $X$ could be, actually corresponds to a *perception.* The perception $P_X$ that a given individual $Y$ has got about how negative or positive a sentence $X$ is. Based on Zadeh's concepts, a fuzzy granulation of positive/negative sentiment using fuzzy intervals is considered to be appropriate to model the problem at hand. Indeed, there are many benefits of being capable of computing the level of intensity of the polarity of a given sentence, as we can determine how strong/weak a given sentiment might be. As a consequence, determining whether the sentiment towards a specific sentence is *moderately* positive/negative, *poorly* positive/negative, *most* positive/negative, etc., is now possible as per the linguistic labels proposed above. Indeed, linguistic polarity intensity could be amenable to be further processed via the *computing with words* methodology introduced by Zadeh [32], making computing with sentiments applications for real data such as products review, possible, as illustrated in Fig. 4.

## 2.6 Calculating Polarity Scores for Sentences

Keep in mind that each word/term in the Sentiment Lexicon we utilise counts with some fundamental attributes. Typically, a lexicon's term will have the following structure:

**(Word Part-of-Speech PositiveScore NegativeScore ObjScore PolarityLabel)**,

where the attribute 'PolarityLabel' can take on three possible values: (i) Positive (pos), (ii) Negative (neg) or (iii) Objective (obj), the latter happening rather sporadically. Hence, if there is a match between a word in a sentence being processed and an entry in the Lexicon (and that match must occur at the 'word name' level and at the 'Part-of-Speech' level), our proposed method *knows* in advance the polarity connotation of the word being analysed, hence it will be in a situation to compute polarities for specific sentences where specific words are present, as it will become evident later on in this Section.

Given a test dataset, the computation of the semantic orientation of sentences requires a two-step approach:

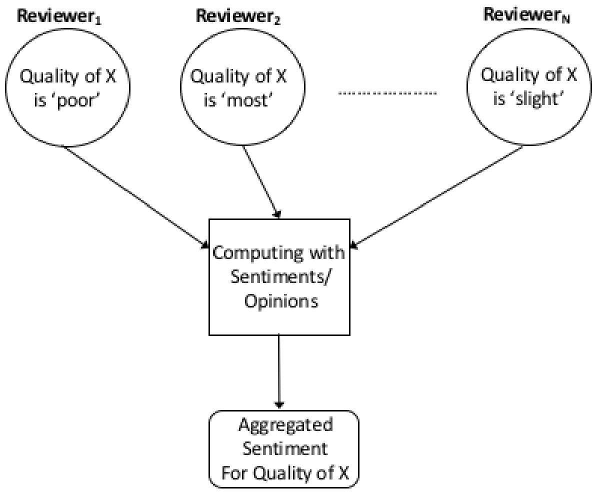



**Fig. 4** Computing with Sentiments - General Diagram

### *2.6.1 Step I: Sentence Polarity Estimation (HSC)*

In the process chain described below, every intermediate step has a list of features as an outcome that is used in the next step.

1. Input = Output of chaining together the Tokeniser, the PoS-tagger and the Smart Parser.
2. Apply Table 1 Semantic Rules.
3. Generate a list of key PoS (adjective, nouns, verbs and adverbs) that convey a sentiment or an opinion.
4. Search the Sentiment Lexicon for the associated matches to substitute the words in the generated list-expression.
5. Generate exception list of key PoS not found associated matches in the Sentiment Lexicon.
6. Calculate each sentence semantic orientation as per the Semantic Rules by considering:
   (a) the word-label semantic orientation present in the Lexicon (POS, NEG or OBJ), and
   (b) the Positive and/or Negative Scores of the words in the sentence that appear in the lexicon.

   Each sentence is assigned a classification label from set {Positive, Negative, Objective, No-Semantic-Orientation}.
7. Use the services of a previously generated dictionary to scan again the resulting list and search for words still labelled as OBJ to be converted into either POS or NEG labels.
8. Generate a new list with all sentences in set dataset and their classification (either positive or negative).

### *2.6.2 Step II: Estimation of the Intensity of the Polarity of a Sentence (HAC)*

This process enhances the Hybrid Standard Classification (HSC) process by:

1. Implementing a fuzzy approach rather than a crisp method to determine the degree with which a given sentence leans towards being *positive* or *negative*:

$$\min\left(IP(Word_1^k)\ \ldots\ IP(Word_n^k)\right) = \delta^k,$$

where $n$ corresponds to the number of words carrying sentiment in a given sentence $S^k$ that are found in the Opinion Lexicon; $IP(Word_i^k)$ is the Intensity Polarity of the $i^{th}$ word $Word_i^k$; and $\delta^k \in [0 \ldots 1]$.

$$\max\Big(\mu_{poor}(\delta^k),\ \mu_{slight}(\delta^k),\ \mu_{moderate}(\delta^k),\ \mu_{very}(\delta^k),\ \mu_{most}(\delta^k)\Big) = \beta^k,$$

Notice that the value of $\beta^k \in [0 \ldots 1]$ is used as the degree with which a given sentence leans towards being *positive* or *negative* but also to identify the classification of the polarity with the corresponding label associated to such degree value $j \in \{poor, slight, moderate, very, most\}$.

2. Diagnosing those sentences with ($\beta^k \approx 0$ that could be considered rather *objective/neutral* as opposed to either *positive* or *negative*.

### *2.6.3 Computation of a sentence semantic orientation (SOR)*

In the process of computing the semantic orientation (SOR of a sentence), the Positive/Negative (pos/neg) label and intensity polarity associated to the sentence words in the lexicon are used. Given a sentence $S_k$ the number of words with positive and negative orientations, respectively, are compared as follows:

**If** *count*(positive words) $>$ *count*(negative words) **then** [the sentence is classified as 'positive'], hence $SOR =$ 'Positive';
**If** *count*(positive words) $<$ *count*(negative words) **then** [the sentence is classified as 'negative'], hence $SOR =$ 'Negative';
**If** *count*(positive words) $=$ *count*(negative words) **then** [There is a tie. Follow alternative process], hence $SOR =$ Table 3 result.

The presence of *ties* is resolved using a mutually exclusive and sequential three-level stratified algorithm as per Table 3). The *count* component above computes how many words with

**Table 3** Stratified algorithm for resolving SOR count ties

| Strata | Task |
|---|---|
| Stratus 1 | The intensity polarity ($IP$) are reviewed and the highest value among negative and positive words wins |
| Stratus 2 | If the previous step fails to produce a classification, apply the hierarchy of importance of PoS particles from most to least *influential*, so that a higher priority is assigned to the $IP$ values of adjectives, followed by adverbs, verbs and nouns |
| Stratus 3 | If the two previous steps fail, examine our dictionary and search for the participant words; extract the frequencies with which each word has appeared in a sentence with a specific polarity (pos/neg); the polarity associated to the highest value wins |

a *positive meaning* are part of a sentence and how many are *negative* (Step 1 Stratified Algorithm). In second place, if Step 1 cannot resolve the problem (there is a tie-break situation), a higher count of words that hold a higher position in the part-of-speech scale would win (i.e. Adjectives being higher in the scale than Nouns). If none of the above Steps 1 and 2 in the Stratified Algorithm work (are not capable of breaking a tie), then we go with the last step (Step 3) of the Stratified Algorithm which will rely on previous occurrences of the word in question, according to the information available in a Word-Frequency Dictionary that has been built in advance. Let us take a look at a specific sentence as an example:

*Example 2* The hotel was nice and apparently clean but some mouse appeared in the room.

Our proposed method will apply the Semantic Rules and will identify that the sentence being handle *contains* the particle 'but' (Rule 11 in the Semantic Rules Table in the article). As per that rule, the semantic of the example sentence corresponding to the sub-sentence occurring before the particle 'but' should be discarded, keeping only the second sub-sentence ('some mouse appeared in the room') in order to determine the semantic orientation. The latter sub-sentence has a clear negative connotation and its semantic orientation will be assigned to the whole sentence being initially analysed ('The hotel was nice and apparently clean but some mouse appeared in the room').

If a sentence $S_k$ is made of $n$ sub-sentences $(S_1, S_2, \ldots, S_n)$, then the *Computed Semantic Orientation* ($CSO$) of $S_k$ is calculated by $SOR$ sub-sentence counting.

1. **If** *count*(Positive $SOR$ Sentences) $>$ *count*(Negative $SOR$ Sentences) **then** $CSO_{(S_1,S_2,\ldots,S_n)}$ = Positive
2. **If** *count*(Positive $SOR$ Sentences) $<$ *count*(Negative $SOR$ Sentences) **then** $CSO_{(S_1,S_2,\ldots,S_n)}$ = Negative
3. **If** *count*(Positive $SOR$ Sentences) = *count*(Negative $SOR$ Sentences) **then** $CSO_{(S_1,S_2,\ldots,S_n)}$ = $SOR$ of $S_k$; $IP(S_k) = \max\{IP(S_1), IP(S_2), \ldots, IP(S_n)\}$

## 3 Experimental Results

This article focuses on SA at the sentence level. Research reported in [27] demonstrates that NB outperforms Support Vector Machines (SVMs) for 'snippets', thus the sentiment/opinion polarity determination comparison between our proposed hybrid method (HSC & HAC) will be done against NB and ME supervised machine learning classification methods.

### 3.1 Datasets description

The first dataset used in our experimental comparison is the *Movie Review Dataset* [20], which has been discussed extensively discussed and used in many SA validation studies to validate [19,21]. The dataset consist of 5,331 of positive and negative snippet, respectively. Each snippet can contain more than one sentence. The second dataset to use as results-validation is *Sentiment140* (available at `http://help.sentiment140.com/for-students`), which consists of twitter data (with emoticons stripped off). This dataset is provided for academic purposes only and can be downloaded in CSV format.

### 3.2 Naïve Bayes Classifier (**NB**)

The NB classifier was trained following the recommendations given in [23]. This classifier uses the concept of 'bag of words' to represent a sentences by creating a 'feature vector' exhibiting the sentence main traits. The NB classifier is a binary classifier, and as such a sentence is classified either as 'negative' or 'positive' by returning a probability value representing the *probability* that such a sentence belongs to a specific label (negative or positive). A sentence is classified to the label with associated probability value of *0.5*. Table 4 shows the results of applying the NB classification algorithm to the test data sets.

### 3.3 Maximum Entropy (**ME**)

The ME classifier was also trained following the recommendations given in [23]. The Generalized Iterative Scaling (GIS) learning method was used to train the ME classifier. As with NB, for a given sentence the ME returns a probability value that represents the *probability*

**Table 4** NB method

| Dataset | Indicator | Value |
|---|---|---|
| Movie | Accuracy | 0.6717 |
| Movie | Precision | 0.6274 |
| Twitter | Accuracy | 0.6785 |
| Twitter | Precision | 0.6315 |

**Table 5** ME method

| Dataset | Indicator | Value |
|---|---|---|
| Movie | Accuracy | 0.6757 |
| Movie | Precision | 0.6291 |
| Twitter | Accuracy | 0.6759 |
| Twitter | Precision | 0.6293 |

**Table 6** HSC method

| Dataset | Indicator | Value |
|---|---|---|
| Movie | Accuracy | 0.7585 |
| Movie | Precision | 0.7278 |
| Twitter | Accuracy | 0.8802 |
| Twitter | Precision | 0.8424 |

**Table 7** HAC Classifier Results for **POS** Dataset - *Increased Granularity*

| | |
|---|---|
| **False Negatives** | 929 |
| **No Semantic Orientation** | 35 |
| **True Positives** | 4,402 |
| **Poor** | 577 |
| **Slight** | 1,106 |
| **Moderate** | 1,041 |
| **Very** | 1,365 |
| **Most** | 313 |
| **Number of Snippets** | 5,331 |

**Table 8** HAC Classifier Results for **NEG** Dataset - *Increased Granularity*

| | |
|---|---|
| **False Positives** | 1,646 |
| **No Semantic Orientation** | 76 |
| **True Negatives** | 3,685 |
| **Poor** | 770 |
| **Slight** | 1,089 |
| **Moderate** | 789 |
| **Very** | 864 |
| **Most** | 173 |
| **Number of Snippets** | 5,331 |

that such a sentence belongs to a specific label (negative or positive). A sentence is classified to the label with associated probability value of *0.5*. Table 5 shows the results of applying the ME classification algorithm to the test data sets.

### 3.4 Hybrid Method (HSC/HAC)

The results achieved by the proposed hybrid HSC correspond to the performance indicators after a second pass of the method, once the missing terms in the sentiment lexicon -if any- have been incorporated (Step I, Section 2.6). They are given in Table 6. The proposed method is clearly superior to the NB and ME supervised machine learning classification methods both in accuracy and in precision and for both datasets.

In addition, when the fuzzy set based approach (HAC) is implemented, it is possible to provide a finer granularity level in the classification process as shown in Table 7 and Table 8 for the Movie Review dataset.

#### *3.4.1 Contribution of methods' sub-components to final outcome*

In this section we will present the contribution of the individual components in the proposed approach to the final outcome. Table 9 reflects how the different components increased the efficiency (in terms of Precision) as they were incrementally added to the proposed solution.

## 4 An alternative approach - Challenges

We have considered whether a different approach could lead to the calculation of both orientation polarity and polarity intensity in one step, instead of two as in the proposed method described in Section 2. In the next paragraphs we will present an alternative strategy that was pursued, but before that, we will provide the theoretical background required in order to describe the alternative method.

It is well known that neither t-norm operators nor t-conorm operators allow "low" values to be compensated by "high" values or viceversa. However, "uninorm operators may allow values separated by their identity element to be aggregated in a compensating way" [9].

| Technique incorporated | Precision (%) | Accumulated Impact (%) |
|---|---|---|
| Using pre-existing semantic rules | 76.77 | |
| Adding effective PoS tagging | 79.33 | 3.33 |
| Adding smart negation handling | 81.17 | 5.73 |
| Adding new semantic rules (R14 & R15) | 83.36 | 8.58 |
| After $2^{nd}$ pass (once the lexicon has learnt new terms) | 84.24 | 9.73 |

**Table 9** Impact of different techniques in hybrid approach precision (Twitter dataset)

Yager and Rybalov [30] provided the following representation of uninorms in terms of a strictly increasing continuous function of a single variable $\phi\colon [0,1] \longrightarrow [-\infty, \infty]$ (generator function):

$$U(x,y) = \phi^{-1}\left[\phi(x) + \phi(y)\right] \;\; \forall x, y \in [0,1]^2 \backslash \{(0,1),(1,0)\}.$$

such that $\phi(0) = -\infty$, $\phi(1) = +\infty$. Chiclana et al. in [5] proved that the and-like representable uninorm operator with $e = 0.5$ and $\phi(x) = \ln \frac{x}{1-x}$ [14], known as the cross-ratio uninorm,

$$U(x,y) = \begin{cases} 0, & (x,y) \in \{(0,1),(1,0)\} \\ \dfrac{xy}{xy + (1-x)(1-y)}, & \text{Otherwise.} \end{cases} \tag{1}$$

is the solution to the functional equation modelling the concept of cardinal consistency of reciprocal preference relations. Fodor [9] extended the cross-ratio uninorm with the identity element $e = 0.5$, so the identity element $e$ can take on any value in $]0,1[$:

$$U(x,y) = \begin{cases} 0, & (x,y) \in \{(0,1),(1,0)\} \\ \dfrac{(1-e)xy}{(1-e)xy + e(1-x)(1-y)}, & \text{Otherwise.} \end{cases} \tag{2}$$

Expression (2) presents the cross-ratio uninorm as an aggregation operator of two arguments. However, associativity property allows uninorm operators to fuse $n$ ($> 2$) arguments:

$$U(x_1, x_2, \ldots, x_n) = \begin{cases} 0, & \text{if } \exists\, i,j : (x_i, x_j) \in \{(0,1),(1,0)\} \\ \dfrac{(1-e)^{n-1}\prod_{i=1}^{n} x_i}{(1-e)^{n-1}\prod_{i=1}^{n} x_i + e^{n-1}\prod_{i=1}^{n}(1-x_i)}, & \text{Otherwise.} \end{cases} \tag{3}$$

Values in the interval $[-1,1]$ could be used as well. Indeed, if we were interested in having semantic orientation values in $[-1,1]$, then according to [25], there is the possibility of using the modified combining function $C\colon [-1,1]^2 \to [-1,1]$ proposed by van Melle [4]:

$$C(x,y) = \begin{cases} x + y(1-x) & , if \min(x,y) \geq 0 \\ x + y(1+x) & , if \max(x,y) \leq 0. \\ \dfrac{x+y}{(1-\min(|x|,|y|)} & \text{,Otherwise.} \end{cases} \tag{4}$$

Notice that $C$ is not defined in the points $(-1,1)$ and $(1,-1)$. However, as per Rudas and Fodor [25], rescaling function $C$ to a binary operator on $[0,1]$, it is possible to obtain a representable uninorm with identity element 0.5 and "as underlying t-norm and t-conorm the product and the probabilistic sum." [25]. This result allows therefore to provide the following definition of $C$ in $(-1,1)$ and $(1,-1)$: $C(-1,1) = C(1,-1) = -1$.

As per the arguments presented above, we propose using the combining function $C(x,y)$ (Eq. 4) as an aggregation of uncertainties in the positive and negative polarity scores of words participating in the computation of the semantic orientation of a given sentence.

## 4.1 Alternative candidate method using Uninorms

Let us describe then the approach that will be followed in performing the calculation of the orientation and intensity values in one step. With regard to the proposed method described in Section 2, the following two components are affected:
- The structure of the sentiment lexicon
- The way that the computing of polarity and its intensity is performed

Let us address both components below.

### *4.1.1 Alternative lexicon structure for the alternative one-step method*

If we recall the structure of the lexicon described in Section 2.1, a Polarity Label (PL) attribute is needed in order to be able to tell whether a given word carried a positive or negative polarity. Notice that to be able to perform computations using any of the equations presented in Section 4, a label in text format is not appropriate as a numerical input is expected. As such, we would have to use the actual Polarity Intensity (PI) as the only element indicating what the semantic orientation of a word is: PSC (Positive Score) and NSC (Negative Score), respectively. As such, we propose a change in the polarity range of words in our sentiment lexicon, introducing the range $[-1, 1] \in \mathbb{R}$. As we must clearly differentiate the values that PSCs and NSCs can take, we have to map to new polarity scores all elements in the sentiment lexicon, as follows:
- Negative scores are mapped from $[0, 1]$ to $[-1, 0]$, with $-1$ meaning *most negative*
- Positive scores are mapped from $[0, 1]$ remaining unchanged, with $+1$ meaning *most positive*
- Zero represent the neutral element

From now on, a calculated negative value points towards a word with negative polarity, whilst a calculated positive score means positive polarity. A computed number $S$ obtained by combining negative and positive scores can be easily identified now depending exclusively on its sign: if $S > 0$ it is positive; if $S < 0$ it is negative, and a value of $S = 0$ implies a neutrality/objectivity case. The PL label initially defined in the sentiment lexicon will not be required any longer.

### *4.1.2 Computing sentence polarity using the alternative one-step method*

As a consequence of the changes in the sentiment lexicon and the utilization of Equation (4), the alternative process to compute both polarity and its intensity is presented as follows:
1. For a given sentence $S$, count all words $x$ with $PSC = 1$ and all words $y$ with with $NSC = -1$
   - If $x > y$ then output a value of 1 (Positive Meaning)
   - If $x < y$ then output a value of $-1$ (Negative Meaning)
   - If $x = y$ then discard all $x$ words with $PSC = 1$ and all $y$ words with $NSC = -1$
   - **If** no sentiment conveying words are left **then** outcome 0 (Neutral Meaning), otherwise go to Step 2.
2. We apply function $C(x, y)$ (Equation 4) only to negative values (the result is stored for future use)
3. We apply function $C(x, y)$ (Equation 4) only to positive values (the result is stored for future use)
4. We apply function $C(x, y)$ (Equation 4) to the results obtained in step 2 and step 3 above. The final result $Rst$ provides us with the final polarity and intensity of the evaluated sentence $S$.
   - If $Rst = 0$, implies Neutral Polarity
   - If $Rst > 0$, implies Positive Polarity, otherwise, that implies Negative Polarity

When the final result $Rst$ satisfies that $Rst \in [-\epsilon, \epsilon]$ with $\epsilon$ representing a value that is very close to 0, then the polarity is neutral/objective, taking a value of 0. The closer $Rst$ is to $+1$ the more *positive* the sentence is; whilst the closer $Rst$ is to $-1$ the more *negative* the sentence is. The same fuzzy sets shown in Section 2.5 could be applied to quantify the intensity of a given polarity.

**Table 10** Comparing results between Hybrid and Alternative uninorm method **- Twitter Dataset**

| Metric | Hybrid Method (Sec. 2) | Alternative Method (Sec. 4) |
|---|---|---|
| **Accuracy** | 0.8802 | 0.7827 |
| **Precision** | 0.8424 | 0.7385 |

**Table 11** Comparing results between Hybrid and Alternative uninorm method **- Movie Dataset**

| Metric | Hybrid Method (Sec. 2) | Alternative Method (Sec. 4) |
|---|---|---|
| **Accuracy** | 0.7585 | 0.6223 |
| **Precision** | 0.7278 | 0.6030 |

## 4.2 Experiments confirming the validity of the initially proposed hybrid method

Experimental results using the alternative method are provided in Tables 9 and 10. Notice that these results are *worse* than those exhibit by the proposed hybrid method (as presented in Sections 2 and 3). In a way, the fact that the alternative method is not better than the initially proposed method (HSC/HAC) confirms that there are a number of tasks executed in the proposed original hybrid method that were properly devised and implemented, which for the Movie Review dataset outperforms the alternative approach in 13.62% in Accuracy and 12.48% in Precision.

## 4.3 Why the polarity scoring method in HSC/HAC works better than the alternative uninorm-based approach

During the analysis of the polarity outputs of HSC/HAC it was noticed that there were multiple situations when there were ties that needed to be resolved. We believe that the main reasons why the HSC/HAC produced better results than the alternative uninorm-based approach are the following:

- The algorithm presented in Table 3 addressed problems in a versatile way as it has the ability to pick the polarity label (pos/neg) of the word with the highest associated polarity in the group of participating words in order to solve a tie.
- This same algorithm has the ability to resort to an established hierarchy among the part-of-speech (PoS) particles in the sentiment-conveying words and pick the label pos/neg linked to the part-of-speech term that is the highest in the hierarchy.
- Lastly but not least, the dictionary of word-occurrence-frequencies that we have built can resolve ties when everything fails, by picking the pos/neg label that has occurred *most often* -highest frequency of occurrence- for a given word.

As opposed to that, the pure direct calculation of the polarity/intensity using *exclusively* the combination function $C(x, y)$ discussed in Section 4 does not have the versatility of the algorithm described in Table 3. An alternative path would be to try to incorporate this diverse process into the calculation of the uninorm values already described, perhaps as exceptions. Another problem we notice is that two different but close values, let's say 0.6231 and 0.6233, would not lead to a tie when compared. However, they are so close together that practically speaking, they do constitute a tie that must be broken. It is in situations like these when the methods discussed in Table 3 may be producing better results as well.

Additionally, in natural languages it is common to use several adjectives stuck together to enforce a given sentiment. For example, "She had a shiny beautiful dark hair". Even if the polarity scores are low for the sentiment-carrying words in the sentence (let's say that 'beautiful' scores high whilst 'shiny' and 'dark' are lower in the polarity value), *still* the power of the adjectives combined together generate a sort of superlative. In this situation, word counting would do better than polarity score amalgamation using an averaging or compensatory combining operator.

## 5 Conclusions

This paper presented a hybrid approach to the SA problem with high levels of *accuracy* and *precision*. It has been proved experimentally that the hybrid SA approach outperforms both the Naïve Bayes (NB) and Maximum Entropy (ME) supervised machine learning classification methods for two datasets widely used in the SA discipline. This result supports the initial hypothesis that a hybrid method using natural language processing techniques, semantic rules and fuzzy sets could be more effective at addressing the challenges of SA than specific techniques in isolation. Additionally, a finer granularity level of sentence polarity is possible with the use of fuzzy sets. A one step alternative approach to compute semantic orientations and their intensities has also been presented and analysed experimentally. However, the results obtained were not better than the original two step process, which was believe to be due to the lack of a mechanism to compensate for the absence of a hierarchy among particles-of-speech, a research avenue that we intend to address in the near future. In closing, there are some lessons learnt and observations that are worth remarking: (a) using efficient NLP techniques (like tokenising, parsing, negation handling, etc.) contribute positively to the application of the proposed Hybrid Method; (b) the creation of an *improved* Sentiment Lexicon is decisive in obtaining good experimental results (SentiWordNet [8] became an important component of the proposed solution and certainly *enriched* dramatically the quality of our Lexicon); (c) work should continue on improving the completeness and quality of Semantic Rules. In essence, hybrid techniques can play an important role in the advancement of the SA discipline by combining together the elements we described in our research contribution.

## References

1. K. S. M. Anbananthen and A. M. H. Elyasir. Evolution of opinion mining. *Australian Journal of Basic and Applied Sciences*, 7(6):359–370, 2013.
2. O. Appel, F. Chiclana, and J. Carter. Main Concepts, State of the Art and Future research Questions in Sentiment Analysis. *Acta Polytechnica Hungarica - Journal of Applied Sciences*, 12(3):87–108, May/June 2015.
3. C. Brenga, A. Celotto, V. Loia, and S. Senatore. *Capturing Digest Emotions by Means of Fuzzy Linguistic Aggregation*, pages 113–139. Springer International Publishing, Cham, 2016.
4. B. Buchanan and E. E. Shortliffe. *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project.* Addison-Wesley, Reading, MA, 1984.
5. F. Chiclana, E. Herrera-Viedma, S. Alonso, and F. Herrera. Cardinal consistency of reciprocal preference relations: A characterization of multiplicative transitivity. *IEEE Transactions on Fuzzy Systems*, 17(1):14–23, 2009.
6. S. R. Das, M. Y. Chen, T. V. Agarwal, C. Brooks, Y. shee Chan, D. Gibson, D. Leinweber, A. Martinez-Jerez, P. Raghubir, S. Rajagopalan, A. Ranade, M. Rubinstein, and P. Tufano. Yahoo! for amazon: Sentiment extraction from small talk on the web. In *8th Asia Pacific Finance Association Annual Conference*, 2001.
7. F. Dzogang, M.-J. Lesot, M. Rifqi, and B. Bouchon-Meunier. Expressions of Graduality for Sentiments Analysis – A Survey. In *Fuzzy Systems (FUZZ), 2010 IEEE International Conference on*, pages 1–7, 2010.
8. A. Esuli and F. Sebastiani. SentiWordNet – A Publicly Available Lexical Resource for Opinion Mining. *In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC06)*, pages 417–422, 2006.
9. J. Fodor. On Rational Uninorms. *In Proceedings of the First Slovakian-Hungarian Joint Symposium on Applied Machine Intelligence, Herlany, Slovakia, February 12-14, 2003*, pages 139–147, 2003.
10. V. Hatzivassiloglou and K. McKeown. Towards the automatic identification of adjectival scales: Clustering adjectives according to meaning. In L. K. Schubert, editor, *ACL: 31st Annual Meeting of the Association for Computational Linguistics, 22-26 June 1993, Ohio State University, Columbus, Ohio, USA, Proceedings*, pages 172–182. ACL, 1993.

11. V. Hatzivassiloglou and K. McKeown. Predicting the semantic orientation of adjectives. In *Proceedings of the 35th Annual Meeting of the ACL and the 8th Conference of the European Chapter of the ACL. New Brunswick, NJ, USA: ACL*, pages 174–181, 1997.
12. M. Hu and B. Liu. Mining and summarizing customer reviews. *Proceedings – ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004 full paper), Seattle, Washington, USA, Aug. 22–25*, 2004.
13. J. Kamps, M. Marx, R. J. Mokken, and M. de Rijke. Using WordNet to measure semantic orientations of adjectives. In *Proceedings of LREC-04, 4th International Conference on Language Resources and Evaluation*, volume IV of *LREC '04*, pages 1115–1118, 2004.
14. E. Klement, R. Mesiar, and E. Pap. On the relationship of associative compensatory operators to triangular norms and conorms. *Int. J. Uncertainty, Fuzziness Knoledge-Based Systems*, 4:129–144, 1996.
15. B. Liu. *Sentiment Analysis and Opinion Mining.* Morgan and Claypool Publishers: Synthesis Lectures on Human Language Technologies, 1st edition, 2012.
16. V. Loia and S. Senatore. A fuzzy-oriented sentic analysis to capture the human emotion in web-based content. *Knowledge-Based Systems*, 58:75–85, Mar 2014.
17. G. Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Rev.*, 63:81–97, 1956.
18. S. Nadali, M. Murad, and R. Kadir. Sentiment classification of customer reviews based on fuzzy logic. *In Information Technology (ITSim), 2010 International Symposium in (Volume:2), Kuala Lumpur, Malaysia*, 2:1037–1040, June 2010.
19. B. Pang and L. Lee. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics (ACL '05)*, ACL'05, pages 115–124, 2005.
20. B. Pang and L. Lee. *Opinion mining and sentiment analysis.* NOW: the essence of knowledge, Foundations and Trends in Information Retrieval, Vol. 2, Nos. 1–2, pp. 1–135, 2008.
21. B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? Sentiment Classification using Machine Learning Techniques. *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 10:79–86, 2002.
22. P. Pérez-Asurmendi and F. Chiclana. Linguistic majorities with difference in support. *Applied Soft Computing*, 18:196–208, 2004.
23. J. Perkins. *Python Text Processing with NLTK 2.0 Cookbook.* Packt Publishing, 2010.
24. C. Potts. Sentiment Symposium Tutorial: Linguistic structure (part of the Sentiment Analysis Symposium held at San Francisco, november 8-9, 2011). Stanford Department of Linguistics, Stanford University. Accessed date: December 2014, November 2011.
25. I. J. Rudas and J. Fodor. Information Aggregation in Intelligent Systems Using Generalized Operators. *International Journal of Computers, Communications & Control*, I(1):47–57, 2006.
26. P. Subasic and A. Huettner. Affect Analysis of Text Using Fuzzy Semantic Typing. *IEEE Transactions on Fuzzy Systems*, 9(4):483–496, August 2001.
27. S. Wang and C. D. Manning. Baselines and Bigrams: Simple, Good Sentiment and Topic Classification. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL 2012): Short Papers*, 2:90–94, 2012.
28. J. Wiebe. Learning subjective adjectives from corpora. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*, pages 735–740. AAAI Press, 2000.
29. Y. Xie, Z. Chen, K. Zhang, Y. Cheng, D. K. Honbo, A. Agrawal, and A. N. Choudhary. MuSES: a multilingual sentiment elicitation system for Social Media Data. *Intelligent Systems, IEEE*, 29(4):34–42, July 2014.
30. R. R. Yager and A. Rybalov. Uninorm aggregation operators. *Fuzzy Sets and Systems*, 80(1):111 – 120, 1996. Fuzzy Modeling.
31. L. A. Zadeh. A new direction in ai: Toward a computational theory of perceptions. *AI Magazine*, 22(1):73–84, 2001.
32. L. A. Zadeh. From Computing with Numbers to Computing with Words - From Manipulation of Measurements to Manipulation of Perceptions. *International Journal of Applied Mathematics and Computer Science (AMCS)*, 12(3):307–324, 2002.